JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1995, **28,** 379–386 NUMBER 3 (FALL 1995)

APPLIED BEHAVIOR ANALYSIS AND STATISTICAL PROCESS CONTROL?

B. L. HOPKINS

AUBURN UNIVERSITY

This paper examines Pfadt and Wheeler's (1995) suggestions that the methods of statistical process control (SPC) be incorporated into applied behavior analysis. The research strategies of SPC are examined and compared to those of applied behavior analysis. I argue that the statistical methods that are a part of SPC would likely reduce applied behavior analysts' intimate contacts with the problems with which they deal and would, therefore, likely yield poor treatment and research decisions. Examples of these kinds of results and decisions are drawn from the cases and data Pfadt and Wheeler present. This paper also describes and clarifies many common misconceptions about SPC, including W. Edwards Deming's involvement in its development, its relationship to total quality management, and its confusion with various other methods designed to detect sources of unwanted variability.

DESCRIPTORS: research methods, applied behavior analysis, statistical process control, total quality management, W. Edwards Deming

---

The research methods of applied behavior analysis have been wonderfully useful. They have allowed us to fruitfully address myriad problems, and to thereby learn which interventions are relatively effective for changing socially important behaviors. In some cases they have even allowed us to become progressively better at addressing certain especially difficult or complex problems (see, e.g., Lovaas, 1993).

The core of our research methodology was borrowed from the experimental analysis of behavior. However, contrasting the current research methods in the *Journal of Applied Behavior Analysis* to those in the *Journal of the Experimental Analysis of Behavior* or noting in Cooper, Heron, and Heward's book (1987) when many of our methods were introduced will reveal considerable evolution in our ways of doing research. Nevertheless, we have developed no clear guides for evaluating new research methodologies. Therefore, responding usefully to suggestions for different research methods is no straightforward matter.

Ideally, we might compare our usual methods to new ones with the primary criterion of acceptance of the new being some improvement in our research. However, we can rarely afford the time or resources to make such comparisons experimentally. Therefore, we must resort to comparing the usual methods and the new ones according to how we predict they would affect our work if we adopted them.

Predicting the likely effects of adopting new methods is problematic. Even trying to make the predictions may border on violation of certain of our principles. We say that, whenever possible, we should decide issues by data, not by argument. When the predicting must be done without data, a reasonable strategy may be to review the relevant, particularly fundamental dimensions of our ways of doing research. Then we would try to predict how the suggested changes in methods would augment or detract from those fundamental dimensions. With the warning that such speculations are risky, I will attempt to evaluate Pfadt and Wheeler's suggestions regarding the potential usefulness of statistical process control (SPC) in contrast to some of the methods we usually use.

The initial drafts of this manuscript were written while the author was on a sabbatical leave supported by Auburn University and hosted by the Department of Psychology, Emory University.

Reprints may be obtained from B. L. Hopkins, Department of Psychology, Auburn University, Auburn, Alabama 36849.

Our work is characterized by interventions that are designed to change behavior. Our baselines and our treatment conditions may be relatively short-lived because we are interested in how the intervention affects the behavior. When we hold conditions constant for long periods of time (see, e.g., Fox, Hopkins, & Anger, 1987), we usually do so after we have intervened and found some procedure to be effective for the problem we are addressing. An important question, then, might be whether the behavior will remain stable during or following the intervention.

SPC methods were originally developed for monitoring important dimensions of manufactured products. However, they can be applied to many phenomena, including questions about whether interventions have important effects, and whether some process is stable or perhaps varying in some unwanted way. Therefore, SPC can be used to address the problems of concern in applied behavior analysis. Predicting whether the use of SPC would likely improve on or detract from applied behavior analysis requires a more detailed examination of the methods involved in the two approaches.

It will be useful to list the methods of SPC so we can compare them to our usual ways of doing research. The interested reader might refer to a standard text on SPC (Grant, 1946). SPC can be used in many different ways. Nevertheless, the following steps are involved in most uses:

1. Some important phenomena are observed.
2. The data that result from the observations are plotted in graphic form.
3. After enough data have accumulated, standard procedures are used to calculate various statistics (e.g., the mean of the data points), with the particular statistic depending on the purposes to which SPC is being put.
4. Further standard statistical procedures are used to calculate levels of the data that would occur infrequently. These levels are called *control limits.*
5. The control limits are plotted as horizontal straight lines on the graph.
6. The data continue to be regularly plotted and, if n ($n \geq 1$; n is typically also defined with statistical methods) successive data points fall outside the control limits, it is assumed that something important is happening with the data; an intervention has an important effect or data are drifting away from a desirable level.

An example will illustrate how these six steps might be carried out. Assume, for example, that a researcher or treatment specialist is interested in how much time an insomniac infant sleeps each day. The researcher might observe and record the daily duration of the infant's sleeping, plot the resulting data in graphic form, calculate the mean daily duration of sleeping, use statistical methods to calculate daily durations of sleeping that would be relatively rare, and plot these rare durations as control limits on the graph. Statistical calculations might suggest that five successive data points falling above the upper control limit or below the lower control limit would be rare. Therefore, if the researcher or applied specialist developed some intervention designed to increase sleeping and, subsequently, five successive data points fell above the upper control limit, it would, following SPC conventions, be interpreted as an indication that the intervention had usefully increased the daily duration of sleeping. Alternatively, if sleeping had been increased for some time to a usefully high level, these graphic and statistical methods could be used to monitor the stability of the high level of sleeping. Then, if five successive data points fell below the lower control limit, it would, following SPC conventions, be interpreted as an indication that something about the conditions yielding the low rate of sleeping had changed so that some further intervention might be appropriate.

I will emphasize two fundamental characteristics of our general research strategies in comparing SPC methods to our usual practices. First, our general research strategy is an interactive one (Johnston & Pennypacker, 1993).

We interact with our subjects or clients through our data that have been transformed into graphic pictures reflecting important dimensions of their behavior. Depending on how our subjects or clients behave, as indicated by our graphs, we wait, intervene, or modify our interventions. Or we may maintain our interventions for an indefinite period of time, if the data indicate that the problem we addressed continues to be satisfactorily improved.

Our subjects and clients affect our behavior by the way they behave. We, in turn, affect their behavior with our interventions. We help them by solving behavior-dependent problems. They educate us about what does and does not produce desired behavior. The graphs of our data allow frequent inspection of the behaviors on which we focus and, if indicated, frequent changes in procedures. The interactive nature of our research is arguably important to our success. We intervene when data indicate it is warranted and maintain our procedures when the data indicate success. Intervening when it is warranted can include changing from an inadequate intervention to one that may be more useful.

SPC, in itself, is not explicitly interactive. However, it advocates frequent observation and accurate data collecting and plotting (the first two steps, above) that are prerequisites for an interactive strategy. In this respect it is compatible with applied behavior analysis and could be used to support an interactive strategy. However, unless a researcher or treatment specialist were well trained in applied behavior analysis, SPC would not provide a usefully complete treatment or research strategy. Therefore, at best, SPC should be considered to be ancillary to applied behavior analysis. This appears to be what Pfadt and Wheeler (1995) are proposing, that SPC be used "for" applied behavior analysis.

The second fundamental characteristic of our research is our intimacy with our subjects' and clients' behaviors (Johnston & Pennypacker, 1993). This intimacy also can easily be seen in our research methods. We spend much time and effort observing the behaviors of interest. In observing these behaviors, we stand back just far enough to ensure that the resulting data are dependable. We sprinkle our research files, our offices, and our journal articles with graphs of important behaviors and their products. By watching our graphs, we maintain intimate relationships with our clients.

As noted above, proponents of SPC advocate frequent accurate observation of important phenomena, graphic displays of the data, and regular inspection of these data displays. In data collection and display, SPC appears to be as supportive of intimacy as applied behavior analysis is.

Other ways in which we try to maintain intimacy with the behaviors we manage and study, however, diverge drastically from what is common in SPC. We have rejected notions such as intrinsic variability and statistical control of variability. We assume that all variability of behavior is a function of natural variables. If control of the variability is important, we search for the environmental sources of variability. This forces us into additional intimate relationships with the behaviors of our subjects or clients, and these relationships may yield further understanding of interactions between environment and behavior. Pfadt and Wheeler (1995) and other proponents of SPC assume that variability within the control limits is acceptable variability. This assumption would likely reduce intimacy.

Pfadt and Wheeler's (1995) example of the woman who had to be restrained and protected from tearing off and swallowing her fingernails or protected from ingesting objects that required surgical removal illustrates how SPC reduces intimacy and, in my opinion, yields apparently poor decisions about interventions. The authors provide SPC analyses of four different kinds of data. The SPC analyses of frequency and duration data yield opinions that the behavior is "stable" and "in statistical control." The analysis of interresponse-time data yields the observation that there was one period of "improved functioning," and the analysis of instantaneous rates yielded the suggestion that "special (unwanted) influences" may have been present on two occasions. Despite the fact that it had been necessary to restrain the woman up to five times per day to keep her from ingesting harmful objects, Pfadt and Wheeler never observe that the treatment was, at best, only partly effective and should have been augmented or changed in some way.

Abandoning a statistical rule for more intimate considerations of what might have been useful and harmful would have likely better served this woman and provided us with an opportunity for possibly learning something about intervening in such cases. The more intimate relationship with this woman's problems might have included considerations of the value of doing something about her problems and a review of potential interventions and their costs. Treatment considerations might reasonably have suggested that some intervention should have been begun after the first day, during which there were four instances in which she had to be restrained for self-injury. If some weighing of treatment and research considerations yielded a decision to postpone an intervention until baseline responding could be better characterized, that intervention would probably have been started sometime between Day 5 and Day 9. Such considerations could have led to help for this woman or to our learning more about treating such cases. SPC yielded the passive judgment, and no intervention for humanitarian or research purposes was recommended during the 30 days for which Pfadt and Wheeler (1995) present data.

It is important to note that we have usually refused to allow relatively automatic processes such as statistics to yield decisions about the importance or long-term effectiveness of an intervention. Our methods again favor a more intimate process in which a researcher or treatment expert looks at many considerations, including the importance of the problem, the value of apparent improvements in the behavior, magnitudes and costs of effects, reliability of the data, trends and variability of the behavior, and the likely effectiveness and costs of alternative treatments.

SPC, in contrast, again promotes the use of methods that are relatively less intimate. The use of statistically derived rules for deciding the importance of an intervention or whether a previously stable performance is deteriorating would again be expected to reduce intimacy.

The fact that some of the statistical calculations require many data points before upper and lower control limits, for example, can be calculated, reduces intimacy in a rarely recognized but potentially very important way. It postpones a decision about the importance of various characteristics of the data until sufficient data have been collected to provide stable statistics. Wheeler and Chambers (1992) provide a useful discussion of this characteristic of SPC. Applied behavior analyses sometimes also require torturously long baselines before decisions about stability of the data are reasonable. However, as a general rule, baselines must be longer to satisfy statistical requirements. In addition, the SPC decision that a baseline is sufficiently long will be based solely on statistical considerations, whereas applied behavior analyses will involve the more intimate considerations described above.

This introduces a very fundamental point about SPC. If a narrow SPC approach is used, what is and is not acceptable is statistically defined. In contrast, the first paper on SPC (Shewart, 1931) described "design specifications" as being different from statistically defined control limits. Design specifications were the practical limits for an important dimension of a manufactured product. Examples might be how large or small a piston could be before a motor wouldn't run or how much water a soup could contain before its taste or texture would be seriously compromised. Shewart explained how the design specifications were more fundamental considerations than the statistically defined limits. So too are our design specifications. If we are to maintain our allegiance to making socially important changes in behaviors (see Baer, Wolf, & Risley, 1968), we must keep social importance, not statistical significance, defining what is and what is not acceptable. Considering social importance forces us into more intimate contact with the variables that would more usefully influence our research and treatment decisions.

It could be argued that Pfadt and Wheeler's (1995) failure to recognize the woman's need for treatment is simply poor SPC rather than a reflection of a fundamental problem with SPC. However, my argument is that following statistical rules rather than engaging in more intimate considerations is likely to yield research and service results like those Pfadt and Wheeler's cases exemplify.

Regarding Pfadt and Wheeler's (1995) example of the 30-year-old man with mental retardation who had "aggressive outbursts," allowing this behavior to continue for 60 days is ethically questionable. That behavior was out of control, in the sense of social importance, on Day 1.

Waiting several days to define control limits statistically will reduce intimacy. Using control limits rather than the design specifications involved in social importance will also reduce intimacy. The reduction in intimacy yields poor treatment and research decisions.

The calculation and use of statistics (the last four steps of SPC described above) would reduce the intimacy between applied behavior analysts and the problems they seek to ameliorate and understand. Some of the examples presented by Pfadt and Wheeler (1995) illustrate this reduction in intimacy and the kinds of unfortunate treatment and research decisions SPC can yield.

In summary, some of the methods of SPC are already a part of applied behavior analysis. The new methods SPC would bring to applied behavior analysis would likely detract from what we already do.

Distinguishing SPC from Other Methods

SPC and some other procedures for promoting product and service quality are often confused. Therefore, it is important to differentiate among them.

SPC was developed and popularized in the 1930s by Walter A. Shewart (1931, 1939). Statistical process control was misnamed. The procedures have nothing to do with controlling anything. Rather, they have to do with methods for statistically analyzing data on the products of some processes. The results of the analyses are most often used with the goal of facilitating decisions about whether the products are being made as wanted or whether the processes producing them may need to be changed in some way. The controlling, itself, involves something other than statistical procedures.

A number of management philosophies and procedures are sometimes associated with SPC. Particularly notable among them today are W. Edwards Deming's views on management as described by total quality management (TQM). In the United States, Deming became something of a legend in the 1980s and early 1990s, and legends are often the subject of misunderstanding. Deming has been credited with inventing SPC and various SPC procedures, first convincing the Japanese to improve the quality of their products, setting the Japanese on the road that led to their supremacy in the quality of manufacturing in several areas, introducing SPC to the Japanese, and developing enormously effective management methods.

Two accounts of some of Deming's work appear to be definitive and are consistent (Mann, 1989; Neave, 1990). The following should be considered no more than a third-hand digest. First, despite some misattributions (see, e.g., Muchinsky, 1993), Shewart, not Deming, invented SPC. Deming, however, was one of the early proponents of SPC. He wrote a number of early books and papers in the field (Deming, 1942, 1944; Deming & Birge, 1934) and is credited with editing Shewart's 1939 paper. Deming took a job with the National Bureau of the Census in 1939 and is credited with many improvements in the efficiency of the Bureau's operations. During World War II, he and other statisticians set up courses to teach American industrialists the methods of SPC as a part of the war effort. This group is credited with some improvements in selected industries. However, Deming felt that the gains made during the war were not sustained after the war, perhaps because there had not been a commitment to SPC by top management.

Deming clearly did not persuade the Japanese, soon after World War II, that their economic future lay in the direction of improved quality of production. The Union of Japanese Scientists and Engineers (JUSE) was formed in 1946 with a broad mission of helping with the industrial reconstruction of Japan. A deserved reputation for poor-quality manufacturing was recognized as an important problem by JUSE members. These members were apparently instrumental in beginning a serious quality movement in the late 1940s and hosted a number of applied statisticians who consulted with them and surely lectured on statistical quality monitoring methods. Meanwhile, Deming was twice invited by General Douglas MacArthur to serve as an advisor to the Japanese census.

During his visits with census personnel, Deming met some of the members of JUSE and was invited to visit them in 1950. Deming visited and lectured in four major cities to managers and engineers and finally asked to meet with top-level managers. He met with the presidents of 21 of Japan's leading companies in a first meeting in the summer of 1950 and lectured to about 500 of them within the next year. By December 1950, the JUSE Board of Directors established the Deming prize to honor Deming's contributions and to recognize an annual recipient organization that made notable efforts to improve the quality of production. Exactly what happened during Deming's first lectures to the top industrialists of Japan is not recorded. However, he obviously began going beyond SPC to talk about broad company commitment to quality and the importance of listening to customers' wishes in all industrial efforts.

During the first three decades following World War II, several Japanese industries came to be the dominant forces in a number of areas such as the manufacturing of automobiles, cameras, and electronics. The high quality of their manufactured goods contributed to their dominance in the relevant markets. The leaders of several industries have credited Deming with being an important force in their efforts to improve the quality of their manufacturing. For example, three pictures hang in the lobby at Toyota's headquarters in Tokyo. One is a picture of the company's founder. A second, of the same size, is of the current chairman of the company. A third and larger picture is of W. Edwards Deming.

By the early 1980s, it was clear that many American manufacturers were in trouble. Many of their products were not as good as those of international companies at the very time that improved communications and transportation were making much of the world one enormous and competitive market. Deming arrived on this scene much more as a teacher of business philosophy and management than as a teacher of statistical process control. His book *Out of the Crisis* (1982) espousing his management views became a big seller, and he spent much of the rest of his long life lecturing and consulting about his views on managing for quality of products and services.

The details of Deming's advice on business philosophy and management have been presented many times and have been discussed in this journal (Mawhinney, 1992; Redmon, 1992). They are a mixture of philosophical statements (e.g., "Create constancy of purpose toward improvement of product and service, with the aim to become competitive and to stay in business, and to provide jobs"), management advice (e.g., "Institute training on the job"), and behavioral nostrums (e.g., "Drive out fear so that everyone may work effectively for the company") (Deming, 1982, p. 23). In addition to Deming's basic rules, he provides a description of "the deadly diseases," such as an emphasis on short-term profit and a lack of constancy of purpose, and the "obstacles" such as "hope for instant pudding," a hope that quality and productivity can be obtained quickly by simple affirmations of intent (Deming, 1982).

In the mid and late 1980s in the U.S., a clear movement developed with Deming as its figurehead but with many other people involved. This movement became a bandwagon, and the general form of the movement became known as TQM. The movement became so popular that a very large percentage of consultants to business must have been forced to say that they knew and could teach other people about TQM. TQM, and Deming's advice on business philosophy, management, and behavior, are often talked about in the same lectures and written about in the same books as SPC. However, they are not SPC.

As far as I know, Deming's advice has never been translated into technologically replicable procedures and tested in reasonably controlled experimentation. There are numerous claims that various companies owe particular successes to Deming's recommendations. For example, see Aguayo's (1990) description of Ford Motor Company's successes soon after Deming began working with them. However, it is unknown what occurred at Ford, what parts of many changes made at Ford were attributable to Deming's advice, and what, if any, effects resulted from Deming's advice. Perhaps the most amazing aspect of TQM and Deming's views is the fact that they became so enormously popular in the absence of hard evidence that they were useful.

As one business fad fades and another becomes stronger, procedures from the older emphasis often appear to become incorporated into the new. Therefore, it sometimes becomes difficult to understand exactly what constitutes the new. This has clearly happened in the case of SPC.

Many other procedures that may have some use for various problems associated with quality have been described as SPC by various authors. Pfadt and Wheeler (1995) mention several of these, for example, flowcharts, Pareto charts, and cause-and-effect diagrams. A flowchart is a simple chart that visually displays the steps in a process. Constructing such a chart might provide some suggestions about where to look for what may be going wrong in a defective product or unsatisfactory service. A Pareto chart is a bar chart that lists possible causes for some result and plots the number of instances that each cause has been suspected to be a culprit for some problem. A Pareto chart yields a ranking of the possible causes so that one might concentrate first on the causes more frequently suspected to be problems. One might first try to fix the most frequently occurring ones. A cause-and-effect diagram lists all the suspected causes of a particular problem and groups them according to categories. These methods are used to try to identify which processes to change when a product or service is not satisfactory. They might be used in conjunction with SPC or applied behavior analysis or any other method for monitoring the data that reflect one or more dimensions of a product or service, but they are not SPC methods.

Conclusions

Statistical process control shares some of the features of applied behavior analysis. The SPC emphasis on data collection and graphing could support an interactive research strategy like that featured in applied behavior analysis. However, in using statistically derived rules for deciding the effectiveness of interventions or the stability of data, SPC would reduce the intimacy that exists between applied behavior analysts and the socially important problems they address. This reduction in intimacy could lead to relatively poor treatment and research decisions. Pfadt and Wheeler (1995) present cases and data that illustrate how following statistical rules results in treatment or research decisions that are probably poor. Because of this reliance on statistics, SPC would likely only detract from applied behavior analysis.

REFERENCES

Aguayo, R. (1990). *Dr. Deming: The American who taught the Japanese about quality.* New York: Carol Publishing.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis.* Columbus, OH: Merrill.

Deming, W. E. (1942). On a classification of the problems of statistical inference. *Journal of the American Statistical Association, 37,* 173–185.

Deming, W. E. (1944). Some principles of the Shewhart methods of quality control. *Mechanical Engineering, 66,* 173–177.

Deming, W. E. (1982). *Out of the crisis.* Cambridge, MA: Massachusetts Institute of Technology.

Deming, W. E., & Birge, R. T. (1934). On the statistical theory of errors. *Reviews of Modern Physics, 6,* 124.

Fox, D. K., Hopkins, B. L., & Anger, W. K. (1987). The long-term effects of a token economy on safety performance in open-pit mining. *Journal of Applied Behavior Analysis, 20,* 215–224.

Grant, E. L. (1946). *Statistical quality control.* New York: McGraw-Hill.

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Erlbaum.

Lovaas, O. I. (1993). The development of a treatment-research project for developmentally disabled and autistic children. *Journal of Applied Behavior Analysis, 26,* 617–630.

Mann, N. R. (1989). *The keys to excellence.* Los Angeles: Prestwick Books.

Mawhinney, T. C. (1992). Total quality management and organizational behavior management: An integration. *Journal of Applied Behavior Analysis, 25,* 525–543.

Muchinsky, P. M. (1993). *Psychology applied to work* (4th ed.). Pacific Grove, CA: Brooks/Cole.

Neave, H. R. (1990). *The Deming dimension.* Knoxville, TN: SPC Press.

Pfadt, A., & Wheeler, D. J. (1995). Using statistical process control to make data-based clinical decisions. *Journal of Applied Behavior Analysis, 28,* 349–370.

Redmon, W. K. (1992). Opportunities for applied behavior analysis in the total quality movement. *Journal of Applied Behavior Analysis, 25,* 545–550.

Shewart, W. A. (1931). *Economic control of quality of manufactured product.* New York: D. Van Nostrand.

Shewart, W. A. (1939). *Statistical method from the viewpoint of quality control.* Washington, DC: Department of Agriculture.

Wheeler, D. J., & Chambers, D. S. (1992). *Understanding statistical process control* (2nd ed.). Knoxville, TN: SPC Press.

Received July 5, 1994
Initial editorial decision April 20, 1995
Revisions received June 2, 1995; June 23, 1995
Final acceptance June 27, 1995
Action Editor, Nancy A. Neef